

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

На правах рукопису

Волощук Орест Романович

УДК 004.853+004.855.5

МЕТОДИ КЛАСТЕРИЗАЦІЇ НА ВЕЛИКИХ МАСИВАХ ДАНИХ

01.05.03 — математичне та програмне забезпечення обчислювальних
машин і систем

Магістерська робота

Науковий керівник

Годич Олесь Васильович,

кандидат фізико-математичних наук, доцент

Львів — 2011

ЗМІСТ

Вступ	3
Розділ 1. Огляд стану проблеми та основні поняття	4

ВСТУП

Актуальність теми. Сьогодні все частіше виникають задачі, так чи інакше пов'язані із розбиттям масиву об'єктів на групи за певними критеріями. Із розвитком обчислювальної техніки збільшуються також об'єми баз даних, що можуть піддаватись такому аналізу. Виникає необхідність створення оптимальних алгоритмів обробки великих масивів даних.

Мета і завдання дослідження. Метою дослідження є розвиток методики кластеризації великих масивів даних. Для досягнення цієї мети були сформульовані та вирішені такі основні завдання:

- провести детальний аналіз та дослідити ефективність різноманітних алгоритмів кластеризації;
- здійснити контроль якості результатів роботи алгоритмів;
- розробити методи та виявити можливі оптимізації, що дозволять прискорити виконання задачі кластеризації

Об'єкт дослідження. Об'єктом дослідження є методи ієрархічні та плоскі алгоритми кластеризації даних.

РОЗДІЛ 1

ОГЛЯД СТАНУ ПРОБЛЕМИ ТА ОСНОВНІ ПОНЯТТЯ

Постановку задачі кластеризації можна знайти ще в "Листі вченому сусіду" Демокріта. Тут він пише, "Якщо тобі, дорогий друже, потрібно розібратись у складному нагромадженні фактів чи речей, ти спершу розклади їх на декілька купок за схожістю. Ситуація проясниться і ти зрозумієш природу цих речей". Сьогодні задачі кластеризації супроводжують більшість задач аналізу даних. початок такої тенденції співпадає із початком розвитку обчислювальної техніки – останній дав змогу застосовувати техніки кластеризації на таких масивах даних, які не піддавались обробці вручну.